

Suchstrategien, Basistechniken, Quellen für die Online-Recherche

RALF PAFFRATH, DFN-GS BERLIN

Abstract

Finding the wanted information in the World Wide Web is becoming a fundamental problem, since the amount of global information that the WWW contains is growing at an incredible rate. In this short paper presented at the GIL-Jahrestagung 98 in Halle (Germany) strategies and different facilities for finding required information in the internet are described. Subsequently an overview for future work in the field of internet information retrieval is given.

1 Einführung

Noch vor einigen Jahren wußten weltbekannte Hard- und Software-Hersteller noch nicht einmal, wie das Wort Internet geschrieben wird. Mit der Einführung des World-Wide-Web-Dienstes begann jedoch der Internet-Boom. Heute gibt es kaum noch ein Unternehmen, welches nicht mit einer Internet-Adresse auf den Visiten-Karten aufwartet. Mit der ständigen Zunahme der Netzteilnehmer (Unternehmen, Hochschulen, Forschungseinrichtungen, Privatnutzer u.a.), nimmt auch die Menge der im Netz vorgehaltenen Informationen zu. Die von vielen Internet-Experten geäußerte Vision, das Internet als Informationsquelle nutzen zu können, ist auf Anhieb nicht so leicht nachvollziehbar. Es gibt im Internet zwar ein unerschöpfliches Angebot an Fakten und Diensten, der Nutzer merkt jedoch aus eigener Erfahrung sehr schnell, wie aufwendig und frustrierend die Informationsbeschaffung sein kann.

Es handelt sich beim vorliegenden Beitrag um eine Zusammenfassung der Ergebnisse eines Vortrages, der während der GIL-Jahrestagung 98 in Halle gehalten wurde. In diesem Vortrag werden dem Nutzer einige praktische und nützliche Tips an die Hand gegeben, mit dem er das Internet als Informationsquelle sinnvoll für sich und seine Zwecke erschließen kann. Nachdem auf die grundsätzliche Problematik bei der Online-Recherche eingegangen wird, werden durch Erläuterungen der Fachtermini, die im Bereich der Online-Recherche allgemein verwendeten Begriffe geklärt. Der mittlere Teil gibt Aufschluß über die unterschiedlichen Suchdienste sowie konkreten Recherchestrategien. Im abschließenden Teil des Vortrages wird auf die zukünftige Entwicklungsrichtung der Suchdienste im Internet eingegangen.

2 Problemstellung

Wie eingangs erwähnt, läutete der World-Wide-Web-Dienst den Siegeszug des Internet ein. Mittlerweile kann man zu fast allen Themen Informationen im Internet bekommen, angefangen z.B. vom "Nußbeckenrezept von Mutter Horn", der Auflösung einiger Zaubertricks von David Copperfield, Bibelzitate, deutschen Gesetzestexten, die Fahrplanauskunft der Deutschen Bahn AG, Informationen zur Giga-Bit-Technologie bis hin zum Quantum Computing usw.¹ Diese Liste kann um zahlreiche Beispiele weitergeführt werden, der Phantasie bei der Themenauswahl sind keine Grenzen gesetzt. Durch die Verwendung weltweit gültiger Standards und einiger weniger zum Teil kostenloser Standardprogramme ist es jedem Netzteilnehmer möglich, eigene Informationen im Netz anzubieten. Als diese Standards entwickelt wurden, hatte jedoch noch

¹Die URLs: <http://www.aixit.com/misc/guildoscorners.htm>; <http://www.snafu.de/tilman/zauber/index.html>;
<http://www.bellnet.com/suchen/kirche/bibel.htm>; <http://www.sunsite.informatik.rwth-aachen.de/germlaws>;
<http://bahn.hafas.de>; <http://www.ifi.uio.no/~sci/papers.html>;
http://www.doc.ic.ac.uk:80/~ids/quantum_computing.html

keiner daran gedacht, mit welcher Wachstumsgeschwindigkeit das weltweite Netz voranschreitet und welche Strukturen für eine gezielte Online-Recherche notwendig sind. Daher ist das heutige Internet alles andere als benutzerfreundlich zu bezeichnen. Folgende grundsätzliche Probleme, die eine Online-Recherche erschweren, lassen sich beschreiben (vgl. ULRICH, B., S. 14-15):

- * Die Informationen, die auf vielen Servern abgelegt werden, sind zum größten Teil unstrukturiert. Texte, Binär-Dateien, wie Audio/Video- und Grafik-Dokumente oder komplette Datenbanken existieren gleichberechtigt nebeneinander. Metadaten (z.B., Autormame, Inhaltsangabe, Erscheinungsjahr) sind selten vorhanden, solche Daten sind jedoch für die gezielte Suche sehr hilfreich.
- * Durch die hohe Eigendynamik unterliegt das weltweite Netz einem ständigen Wandel. Informationen, die man gestern noch unter einer festen Adresse vorgefunden hat, können heute verändert, gelöscht oder verschoben worden sein. Daher ist eine URL-Adresse als Quelle mit Vorsicht anzugeben.
- * Ob eine gesuchte Information im Internet gefunden werden kann, hängt in gewisser Weise auch davon ab, wie häufig auf diese Informationen über einen geeigneten Link verwiesen wird.

Durch Implementierung neuer Standards (z.B. durch SGML, Z39.50)² während der letzten Jahre wurde versucht, die oben beschriebene Problematik anzugehen, aber diese Systeme konnten bisher nicht in die Breite getragen werden. Im Jahre 1994 erschienen die ersten Suchhilfen, die versuchen die endlos erscheinende Informationsmenge zu indexieren. Mittlerweile haben diese Suchhilfen in ihrer Anzahl und Vielfalt zugenommen. Das folgenden Kapitel gibt einen Überblick der wichtigsten Suchdienste.

3 Terminologie, Typologie der Suchhilfen im Internet

Eine einheitliche Terminologie im Bereich der Online-Recherche existiert bisher nicht. So können gleiche Begriffe in ihrer Bedeutung in unterschiedlichen Kontexten stark voneinander abweichen. Eine gelungene analytische Abgrenzung der unterschiedlichen Begriffe wird in (TRAUGOTT, K.: Kapitel Terminologie) vorgenommen. Hier wird zwischen drei Methoden der Navigation unterschieden (vgl. TRAUGOTT, K.: Kapitel Terminologie):

1. Browsing bezeichnet das systematische "Suchen/Blättern" unter Zuhilfenahme von äußeren Strukturen, etwa vergleichbar mit dem kapitelseitigen "Durchblättern" eines Buches.
 2. Surfen bezeichnet die unsystematische Vorgehensweise bei der Informationsbeschaffung, welches mit dem "Durchblättern" eines gesamten Buches vergleichbar ist.
 3. Suchen bezeichnet die gezielte Online-Recherche in einer Datenbank, die den Zugriff auf ein Dokument mittels umfangreicher, zum Teil komplizierter Sucheingaben und Suchbedingungen in einem Schritt ermöglicht.
- Die eigentliche Online-Recherche kann als "iterativer Suchprozeß" (vgl. TRAUGOTT, K.: Kapitel Suchprozeß) bezeichnet werden und setzt sich in der Regel aus der Kombination der oben beschriebenen Vorgänge (Browsing, Suchen, Surfen) zusammen. Je nach Ziel und Strategie, die

² siehe <http://wombat.doc.ic.ac.uk/foldoc/index.html>

beim Suchen verfolgt werden, kann die Recherche mehrere Zyklen durchlaufen, bis sich der Nutzer an das gesuchte Dokument so nah wie möglich herangetastet hat.

Gerade im Hinblick auf die Strategie, die bei der Recherche ausgewählt wird, ist es von Bedeutung, sich der unterschiedlichen Suchhilfen im Internet bewußt zu sein. Das Angebot der Suchhilfen im weltweiten Netz ist so reich und vielfältig, daß eine gute Übersicht erforderlich ist. Eine relativ aktuelle und praktisch brauchbare Übersicht (typologisch, thematisch sortiert) findet der Internetnutzer unter <http://www.beaucoup.com/enggerm.html>. Hier werden die unterschiedlichen Suchhilfen in Allgemeine Suchmaschinen, "Mehrfachsuchmaschinen", "Geographische Spezifizierungen", "Software", "Literaturverzeichnisse/Bildungswesen/Sprachen/Literatur", "Kunst/Musik/Graphik", "Naturwissenschaften/Umwelt/Technologie", "Email/Domains/Telefonnummern/Organisationen", "Computer/Internet/WWW", "Politik/Regierung", "Gesundheit/Medizin/Ernährung" und "Potpourri" gegliedert.

Grundsätzlich kann man zwischen zwei in ihrer Funktionsweise unterschiedlichen Suchhilfen unterscheiden (vgl. ULRICH, B.: S. 45):

1. Die "Suchmaschinen", die versuchen die Internetressourcen komplett zu indexieren und dem Nutzer die Möglichkeit bieten, im Rahmen einer Volltextsuche existierende WWW-Seiten auf bestimmte Begriffe und deren Kombination hin zu untersuchen.
2. Die "thematischen Verzeichnisse" (Kataloge) bieten dem Nutzer eine sinnvoll geordnete Übersicht von Informationsquellen an. Eine komplette Erfassung alle Internetressourcen wird hierbei jedoch nicht angestrebt.

Nachdem in der Kürze auf die unterschiedlichen Typen eingegangen wurde, wird im folgenden Kapitel einige Strategien für die Online Recherche erläutert werden.

4 Recherchestrategien

Eine allumfassende Strategie für die Online-Recherche im weltweiten Netz gibt es nicht. Jeder Internetnutzer sollte sich vor jeder Informations-Recherche fragen, was er sucht (Personen, Firmen, Software usw.) und ob er die gewünschten Informationen nicht schneller über traditionelle Formen der Informationsbeschaffung einholen kann, z.B. via Telefon, Fax, Arbeitskolleg(inn)en usw. Entschließt sich der Benutzer für die Suche im Netz, so sollten folgende grundsätzliche Überlegungen angestellt werden:

1. Wo könnte die gesuchte Information im Netz plaziert sein?
2. Wer könnte die Information im Netz anbieten?
3. Wer könnte ein Interesse an der Verbreitung der gesuchten Informationen haben?

Neben diesen grundsätzlichen Überlegungen ist auch die richtige Wahl der Suchbegriffe bei der Verwendung einschlägiger "Suchmaschinen" (wie z.B. AltaVista, Lycos, HotBoot usw.) von entscheidender Bedeutung. Folgende Faustregel ist dabei zu berücksichtigen: Je allgemeiner der Begriff, desto größer die Anzahl der möglichen Treffer. Bei der Auswahl der Suchbegriffe sollte man also spezielle Begriffe verwenden.

Einige Suchmaschinen bieten auch die sogenannte "Phrasensuche" an. Hierbei hat der Nutzer die Möglichkeit, nach Satzphrasen, bestehend aus mehreren syntaktisch zusammenhängenden Begriffen und Bezeichnungen, zu suchen. Der Benutzer bekommt nur die Dokumente angeboten, in denen die gesuchte Satzphrase vorkommt.

Werden in einer Recherche "thematische Verzeichnisse" verwendet, ist darauf zu achten, daß diese Suchhilfen in der Regel keine einheitlichen Kategorieinteilungen verwenden. Z. B. kann es

vorkommen, daß der Begriff "Börsenkurse" in einem Verzeichnis unter der Kategorie "Infoserver" zu finden ist und in einem anderen Verzeichnis der Kategorie "Wirtschaft" zugeordnet wird.

Viele Suchmaschinen bieten mittlerweile auch die Stichwortsuche in thematischen Verzeichnissen an. Der Benutzer bekommt nach Eingabe eines Stichwortes eine Übersicht der möglichen Kategorien angezeigt, die er dann gegebenenfalls verfeinern kann.

Mit den aktuell verwendbaren Suchhilfen im Internet läßt sich jedoch selbst bei bester Strategie nicht alles finden. Die Grenzen und die zukünftigen Entwicklungsrichtungen der Suchhilfen werden im nächsten Kapitel kurz beschrieben.

5 Grenzen der Suchmaschinen und zukünftige Entwicklungsrichtungen

Eine Suchmaschine kann grundsätzlich nur solche Informationen in ihre Datenbank aufnehmen, welche sie zum Zeitpunkt der Erstellung der Datenbank im Netz vorgefunden hat. Durch das rasche Informationswachstum im Internet, ist es den Suchmaschinen nicht möglich, durch das Verfolgen von Verweisen neue Dokument sofort zu finden und zu erfassen. Daher dauert es eine gewisse Zeit, bis auch "neue" Dokumente mit Suchmaschinen gefunden werden können.

Ein weiteres Problem ist der inkonsistente Datenbestand der Suchmaschinen, der durch das Verschieben und Löschen von Dokumenten hervorgerufen wird.

Abschließend sollen zwei vielversprechende Ansätze, die die zukünftige Landschaft im Internet entscheidend verändern können, kurz vorgestellt werden. Zum einen sind das die "Intelligenten Agenten" und zum anderen die Suchhilfen, die den Datenbestand in visueller Form für den Benutzer erschließen. Die "Intelligenten Agenten" bieten dem Nutzer die Möglichkeit eine Anfrage in natürlich sprachlicher Form an das System zu stellen. Der Agent bearbeitet dann die Anfrage im Hintergrund. Sobald er fündig wird, tritt der Agent wieder in Interaktion mit dem Benutzer und klärt, ob eine weitere Recherche notwendig ist oder ob das bis dahin erzielte Ergebnis aus Sicht des Nutzers ausreichend ist. Ein konkreter Prototyp, ein sogenannter "Internet Fish", wird in (B.A. LAMACCHIA) beschrieben. Die "visuellen Formen" der Informationsdarstellung im WWW soll dem Benutzer die gefundenen Dokumente in 3D Graphiken anbieten. Der Nutzer hat die Möglichkeit, die Informationen räumlich zu erfassen. Dieser Ansatz wird u.a. in (S.JEROMY CARRIERE, RICK KAZMAN) beschrieben.

6 Literatur

- S. JEROMY CARRIERE, RICK KAZMAN (1997): WebQuery: searching and visualizing the Web through connectivity: In Computer Networks and ISDN p. 1257, Verlag Elsevier
- ULRICH, B. (1997): Effektive Suche im Internet. Verlag O'Reilly
- TRAUGOTT, K. (1996): Suchmaschinen im Internet: <http://www.ub2lu.se/tk/demos/DO9603-meng.html>
- B. A., LAMACCHIA (1997): The Internet Fish Construction Kit: In Computer Networks and ISDN p. 1237, Verlag Elsevier